Improving Visual Matching

Michael S. Lew   Nicu Sebe
Leiden Institute of Advanced Computer
Science, Leiden, The Netherlands
{mlew, nicu}@liacs.nl

Thomas S. Huang
Beckman Institute, University of Illinois
at Urbana-Champaign, USA
huang@ifp.uiuc.edu

Abstract

Many visual matching algorithms can be described in terms of the features and the inter-feature distance or metric. The most commonly used metric is the sum of squared differences (SSD), which is valid from a maximum likelihood perspective when the real noise distribution is Gaussian. Based on real noise distributions measured from international test sets, we have found experimentally that the Gaussian noise distribution assumption is often invalid. This implies that other metrics, which have distributions closer to the real noise distribution, should be used. In this paper we considered two different visual matching applications: content-based retrieval in image databases and stereo matching. Towards broadening the results, we also implemented several sophisticated algorithms from the research literature. In each algorithm we compared the efficacy of the SSD metric, the SAD (sum of the absolute differences) metric, the Cauchy metric, and the Kullback relative information. Furthermore, in the case where sufficient training data is available, we discussed and experimentally tested a new metric based directly on the real noise distribution, which we denoted the maximum likelihood metric.

1. Introduction

At the core of many algorithms in computer vision is the metric or similarity measure used to determine the distance between two features. The SSD (sum of the squared differences) and SAD (sum of the absolute differences) are the most commonly used metrics. This brings to mind several questions. First, under what conditions should one use the SSD versus the SAD? From a maximum likelihood perspective, it is well known that the SSD is justified when the additive noise distribution is Gaussian. The SAD is justified when the additive noise distribution is Exponential (double or two-sided exponential). Therefore, one can determine which metric to use by checking if the real noise distribution is closer to the Gaussian or the Exponential. This leads to the second question: What distance measure do we use in comparing the real noise

distribution to the best fit Gaussian or Exponential distributions? This is not an easy question to answer because the choice of the distance measure will bias the comparison. Ideally, we would also like to use a maximum likelihood distance measure to compare the two distributions. However, we would need to measure distributions from a statistically large number of datasets. For example, to obtain one representative distribution, ? , we need a dataset which contains a statistically large number of examples. To find a distribution of the variance of each element of ? , we would need a statistically large number of datasets. In practice, the Chi-square test is frequently used, and since we have not found a better solution, we used it for comparing the distributions.

The common assumption is that the real noise distribution should fit either the Gaussian or the Exponential, but what if this assumption is invalid? What if there is another distribution which fits the real noise distribution better than the Gaussian or the Exponential? It is precisely this question which we examined in this paper. Toward answering this question, we have endeavored to use international test sets and promising algorithms from the color indexing and stereo matching research literature.

Color indexing is one of the most prevalent retrieval methods in content based image retrieval. Given a query image, the goal is to retrieve all the images whose color compositions are similar to the color composition of the query image. Typically, the color content is described using a histogram [15]. In general, color histograms are computed and the histogram intersection criterion is used to compare them. Hafner, et al. [8] suggests the usage of a more sophisticated quadratic form of distance measure, which tries to capture the perceptual similarity between any two colors. In all of these works, most of the attention has been focussed on the color model with little or no consideration of the noise models.

Stereo matching implies finding correspondences between two or more images. If these correspondences can be found accurately and the camera geometry is

known, then a 3D model of the environment can be reconstructed [2]. In [6], pixel correspondences are found by adaptive, multi-window template matching. The templates are compared using the SSD. Recent research by [3] concluded that the SSD is sensitive to outliers and therefore robust M-estimators should be used regarding stereo matching. However, the authors [3] did not consider metrics based on similarity distributions. They considered ordinal metrics, where an ordinal metric is based on relative ordering of intensity values in windows - rank permutations. Cox, et al. [5] presented a stereo algorithm which optimizes a maximum likelihood cost function. This function assumes that corresponding features in the left and right images are normally distributed about a common true value. However, the authors [5] noticed the normal distribution assumption used to compare corresponding intensity values is violated for some of their test sets. They altered the stereo pair so that the noise distribution would be closer to a Gaussian. In our approach, we attempt to find a better model for the real noise distribution instead of altering the stereo pair.

Boie and Cox [4] consider a model of camera noise comprised of stationary direction-dependent electronic noises combined with fluctuations due to signal statistics. These fluctuations enter as a multiplicative noise and are non-stationary and vary over the scene. A substantial simplification appears if the noise can be modeled as Gaussian distributed and stationary. This work is complementary to ours. They try to model the imaging noise. We try to model the noise between two images which are different due to varying orientation, or printer noise.

Section 2 describes the mathematical support for the maximum likelihood approach. The setup of our experiments is given in Section 3. In Section 4 we apply the theoretical results from Section 2 to determine the influence of the real noise model on the accuracy of retrieval methods in color image databases. In Section 5 we study the real noise model to be chosen in stereo matching applications. Conclusions are given in Section 6.

2. Maximum Likelihood Approach

Consider M image pairs (or more generally, feature vectors) from the database (D): $(x_i, y_i) \in D$, with $i = 1, \cdots, M$. The images in each pair were chosen to be similar according to the ground truth (G):

$$x_i \equiv y_i, \qquad i = 1, \cdots, M \qquad (1)$$

Equation (1) can be further written as:

$$x_i = y_i + n_i, \qquad i = 1, \cdots, M \qquad (2)$$

where n_i represent the "noise" image obtained as the difference between the other two images.

In this context, the similarity probability can be defined:

$$P(G) = \prod_{i=1}^{M} \{\exp[-\rho(x_i, y_i)]\} \qquad (3)$$

where function ρ is the negative logarithm of the probability density of the noise.

According to (3), we have to find the probability density function of the noise which maximizes the similarity probability: maximum likelihood estimate for the noise distribution [11].

Taking the logarithm of (3) and using (2) we find that we have to minimize the expression:

$$\sum_{i=1}^{M} \rho(n_i) \qquad (4)$$

where $n_i = x_i - y_i$ and the operation "-" denotes the difference between corresponding elements in the images or in their feature vectors.

To analyze the behavior of the estimate we take the approach described in [9] based on the influence function. The influence function characterizes the bias that a particular measurement has on the solution and is proportional to the derivative, ψ, of the estimate:

$$\psi(z) \equiv \frac{d\rho(z)}{dz} \qquad (5)$$

In the case where the noise is Gaussian distributed:

$$Prob\{x_i - y_i\} \sim \exp(-(x_i - y_i)^2) \qquad (6)$$

then

$$\rho(z) = z^2 \qquad \psi(z) = z \qquad (7)$$

If the errors are Exponential distributed, namely

$$Prob\{x_i - y_i\} \sim \exp(-|x_i - y_i|) \qquad (8)$$

then,

$$\rho(z) = |z| \qquad \psi(z) = sgn(z) \qquad (9)$$

In this case, using (4), we minimize the mean absolute deviation, rather than the mean square deviation. Here the tails of the distribution, although exponentially decreasing, are asymptotically much larger than any corresponding Gaussian.

A distribution with even more extensive tails is the Cauchy distribution,

$$Prob\{x_i - y_i\} \sim \frac{1}{\mathbf{a}^2 + (x_i - y_i)^2} \qquad (10)$$

where $\mathbf{a}$ is a parameter which determines the height and the tails of the distribution.

This implies

$$\rho(z) = \log(\mathbf{a}^2 + z^2) \qquad \psi(z) = \frac{z}{\mathbf{a}^2 + z^2} \qquad (11)$$

For normally distributed errors, (7) says that the more deviant the points, the greater the weight. By
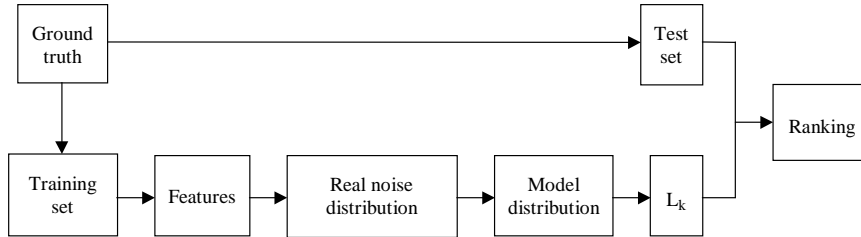
Figure 1. An overview of a similarity matching algorithm

contrast, when tails are somewhat more prominent, then (9) says that all deviant points get the same relative weight, with only the sign information used. Finally, when the tail are even larger, (11) says ψ increases with deviation, then starts decreasing, so that the true outliers are not counted at all.

The additive noise model in (2) is the dominant model used in computer vision regarding maximum likelihood estimate. For example, Haralick and Shapiro [10] consider this model in defining the M-estimate: "... of the form $\min \sum_i \rho(x_i - T_k)$ is called an M-estimate." Note that the operation "-" between the estimate (T_k) and the real data (x_i) implies an additive model.

In summation, one can note that (7) resembles the L_2 metric, while (9) and (11) resemble the L_1 and L_c metrics, respectively. Thus, the maximum likelihood approach gives a direct connection between the noise distribution and the comparison metrics. If ρ is the negative logarithm of the probability density of the noise, then the corresponding metric is given by (4). In practice, the probability density of the noise can be estimated from the normalized histogram of the absolute differences.

3. Experimental Setup

First, we assume that representative ground truth is provided. The ground truth is split into two non-overlapping sets: the training set and the test set as shown in Figure 1. Note that L_k is a notation for all possible metrics that can be used, e.g. L_1, L_2, L_c. Second, the training set is converted to a histogram which is then normalized and denoted as the real noise distribution. The Gaussian, Exponential, and Cauchy distributions are fitted to the real noise distribution. The Chi-square test is used to find the fit between each of the model distributions and the real distribution. We select the model distribution which has the best fit and its corresponding metric (L_k) is used in ranking. The ranking is done using only the test set. For benchmarking purposes in all of the experiments we compare our results with the ones obtained using the Kullback relative information (K) [12]. We chose the Kullback rela-

tive information because it is the most frequently used similarity measure in information theory. Furthermore, Rissanen [14] showed that it serves as the foundation for other minimum description length measures such as Akaike's [1] information criterion. Regarding the relationship between the Kullback relative information and the maximum likelihood approach, Akaike [1] showed that maximizing the expected log likelihood ratio in maximum likelihood estimation is equivalent to maximizing the Kullback relative information.

It is important to note that for real applications, the parameter in the Cauchy distribution is found when fitting this distribution to the real distribution from the training set. This parameter setting would be used for the test set and any future comparisons in that application.

For our image retrieval experiments we considered the applications of color image retrieval in printer-scanner copy location and object recognition by color invariance. In the copy location application, the goal is to find copies of an image taken from a magazine, newspaper, or book. This task involves noise due to the dithering patterns of the printer and scanner noise. Furthermore, it is easy to verify that color printers do not produce the same colors, brightness, or contrast as the original. In object recognition, multiple pictures are taken of a single object at different orientations. Therefore, the correct match for an image is known by the creator of the ground truth.

In stereo matching, the ground truth is typically generated manually. A set of reference points are defined in the images and then a person finds the correspondences for the stereo pair.

In summary, our algorithm for choosing an analytic metric can be described as follows:

Step 1 Compute the feature vectors from the training set

Step 2 Compute the real noise distribution from the differences between corresponding elements of the feature vectors

Step 3 Compare each of the model distributions to the real noise distribution using the Chi-square test

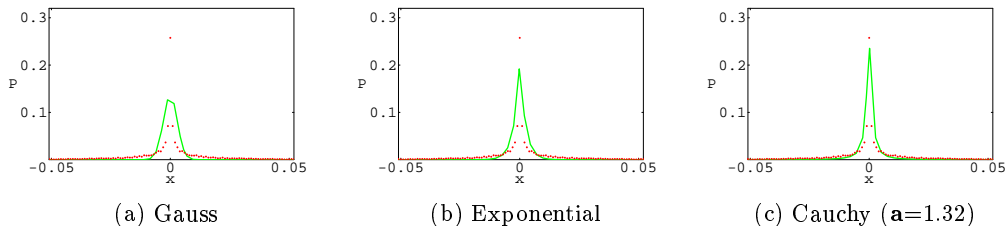|		
(a) Gauss	(b) Exponential	(c) Cauchy (**a**=1.32)

Figure 2. Noise distribution in the Corel database compared with the best fit Gaussian (a) (approximation error is 0.106), best fit Exponential (b) (approximation error is 0.082) and best fit Cauchy (c) (approximation error is 0.068)

Step 3.1 For a parameterized metric such as L_c compute the value **a** of the parameter that minimizes the Chi-square test

Step 4 Select the corresponding L_k of the best fit model distribution

Step 4.1 Use the value **a** found from **Step 3.1** in the parameterized metrics

Step 5 Apply the L_k metric in ranking

As noted in the previous section, it is also possible to create a metric based on the real noise distribution using maximum likelihood theory. Consequently, we denote the maximum likelihood (ML) metric as (4) where ρ is the negative logarithm of the normalized histogram of the absolute differences from the training set. Note that the histogram of the absolute differences is normalized to have area equal to one by dividing the histogram by the total number of examples in the training set. This normalized histogram is our approximation for the probability density function.

4. Similarity Noise in Color Indexing

Our formulation of the image retrieval problem is as follows: Let $\mathcal{D}$ be an image database and $\mathcal{Q}$ be the query image. Obtain a permutation of the images in $\mathcal{D}$ based on $\mathcal{Q}$, i.e assign $rank(\mathcal{I}) \in [|\mathcal{D}|]$ for each $\mathcal{I} \in \mathcal{D}$, using some notion of similarity to $\mathcal{Q}$. The problem is usually solved by sorting the images $\mathcal{Q}' \in \mathcal{D}$ according to $|f(\mathcal{Q}') - f(\mathcal{Q})|$, where $f(\cdot)$ is a function computing feature vectors of images and $|\cdot|$ is some distance measure defined on feature vectors.

We applied the theoretical results described in Section 2 in two experiments. We determined the influence of the similarity noise model on finding similar images which differ due to either printer-scanner noise or change of viewpoint. We used two color image databases. The first one was the Corel Photo database and the second one consisted of 500 reference images of domestic objects, tools, food cans, art artifacts, etc.

For benchmarking purposes in both experiments with color databases we compared our results with the ones obtained using the quadratic distance measure (L_q) proposed by Hafner, et al. [8].

The performance evaluation was formulated as follows: Let $\mathcal{Q}_1, \cdots, \mathcal{Q}_n$ be the query images and for the i-th query $\mathcal{Q}_i$, $\mathcal{I}_1^{(i)}, \cdots, \mathcal{I}_m^{(i)}$ be the images similar with $\mathcal{Q}_i$ according to the ground truth. The retrieval method will return this set of answers with various ranks. As an evaluation measure of the performance of the retrieval method we used recall vs. precision at different scopes: For a query $\mathcal{Q}_i$ and a scope $s > 0$, the recall r is defined as $|\{\mathcal{I}_j^{(i)}|rank(\mathcal{I}_j^{(i)}) \leq s\}|/m$, and the precision p is defined as $|\{\mathcal{I}_j^{(i)}|rank(\mathcal{I}_j^{(i)}) \leq s\}|/s$.

4.1. Experiments with Noisy Copy Pairs

The first experiments were done using 8,200 images from the Corel database. We used this database because it represents a widely used set of photos by both amateur and professional graphical designers. Furthermore, it is available on the Web at http://www.corel.com.

Before we can measure the accuracy of particular methods, we first had to find a challenging and objective ground truth for our tests. The idea of our experiments was to measure the effectiveness of a retrieval method when trying to find a copy of an image in a magazine or newspaper. In order to create the ground truth we printed 82 images using an Epson Stylus 800 color printer at 720 dots/inch and then scanned each of them at 400 pixels/inch using an HP IIci color scanner. Note that we purposely chose a hard test set. The query image is typically very different from the target image. The copy pairs typically differ by color shifts, quantization artifacts, and dithering noise.

We used the HSV color model and quantized H using 4 bits, S using 2 bits and V using 2 bits. The first question we asked was, "Which distribution is a good approximation for the real color model noise?" To answer this we needed to measure the noise with respect to the color model. The real noise distribution was obtained as the normalized histogram of differences between the elements of color histograms correspond-

ing to copy-pair images from the training set (50 image pairs).

In fitting the Exponential, Gaussian, and Cauchy distributions to the real noise distribution, the Cauchy had the best fit followed by the Exponential and then the Gaussian as shown in Figure 2. Consequently, this implies that the Cauchy metric should have the best accuracy followed by the Exponential and Gaussian. From the tests, as shown in Figure 3, it is clear that L_c gives a significant improvement in accuracy as compared to L_2, L_1, and L_q. The Kullback relative information and L_q give better accuracy than L_2 or L_1. Overall, the ML metric gives the best accuracy.

Another interesting performance evaluation is to display the percentage of correct copies found within the top n matches. These results are shown in Table 1.

Top	20	40	100
L_2	48.78	54.87	67.07
L_1	62.19	68.29	84.14
L_q	66.34	73.66	88.29
K	68.29	75.60	86.58
L_c **a**=1.32	71.95	79.26	92.68
ML	75.60	82.92	96.34

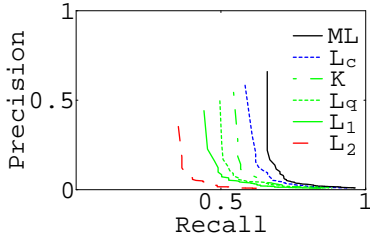**Table 1. Retrieval accuracy in the Corel database**

Figure 3. Precision/Recall in Corel database; for L_c, a=1.32

4.2. Experiments with Object Recognition

In the second experiment we used a database consisting of 500 images of color objects such as domestic objects, tools, toys, food cans, etc. As ground truth we used 48 images of 8 objects taken from different camera viewpoints (6 images for a single object). For this experiment we chose to implement a method designed for indexing by color invariants. Our goal was to study the influence of the similarity noise on the retrieval results.

Gevers, et al. [7] analyzed and evaluated various color features for the purpose of image retrieval by color-metric histogram matching under varying illumination environments. They introduced a new color model l and showed that it is invariant for both matte and shiny surfaces:

$$l_1(R,G,B) = \frac{(R-G)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \quad (12)$$

$$l_2(R,G,B) = \frac{(R-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \quad (13)$$

$$l_3(R,G,B) = \frac{(G-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \quad (14)$$

where R, G, B are the color values in the RGB color space.

The authors [7] concluded that this color model is the most appropriate color model to be used for image retrieval by color-metric histogram matching under the constraint of a white illumination source. This conclusion was drawn using histogram intersection (L_1) as the comparison metric between the color histograms.

Scope	Precision			Recall		
	5	10	25	5	10	25
L_2	0.425	0.258	0.128	0.425	0.517	0.642
L_1	0.45	0.271	0.135	0.45	0.542	0.675
L_q	0.46	0.280	0.143	0.46	0.561	0.707
K	0.466	0.279	0.138	0.466	0.558	0.692
L_c	0.525	0.296	0.146	0.525	0.592	0.733
ML	0.533	0.304	0.149	0.533	0.618	0.758

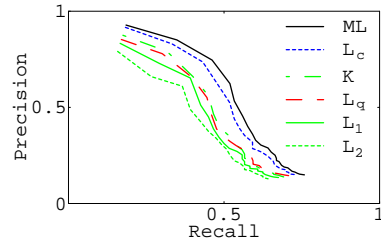**Table 2. Precision/Recall at various scopes for the color objects database; for L_c, a=2.88**

Figure 5. Precision/Recall for the color objects database; for L_c, a=2.88

Using 24 images with varying viewpoint as the training set, we calculated the real noise distribution and studied the influence of different distance measures on the retrieval results. We used the l color model introduced before and we quantized each color component with 3 bits resulting in color histograms with 512 bins.

The Cauchy distribution was the best match for the real noise distribution. The Exponential distribution was a better match than the Gaussian (Figure 4). Table 2 shows the precision and recall values at various scopes. The results obtained with L_c were consistently better than the ones obtained with the other measures. Figure 5 shows the precision-recall graphs. The curve

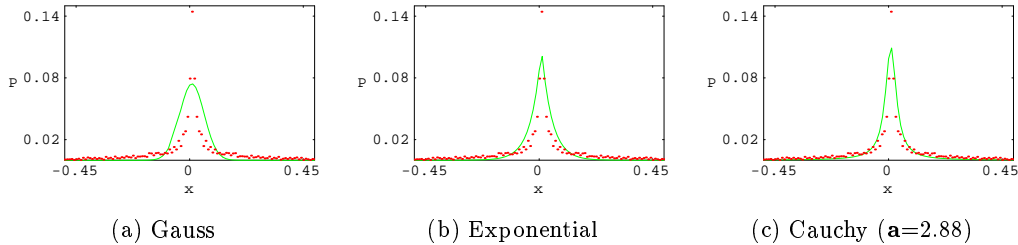(a) Gauss (b) Exponential (c) Cauchy ($\mathbf{a}$=2.88)

Figure 4. Noise distribution in color objects database compared with the best fit Gaussian (a) (approximation error is 0.123), best fit Exponential (b) (approximation error is 0.088) and best fit Cauchy (c) (approximation error is 0.077)

corresponding to L_c is above the others showing that the method using L_c is more effective. Note that the Kullback relative information and L_q perform better than L_1 or L_2.

In summary, L_c performed better than the analytic distance measures, and the ML metric performed best overall. It is interesting that the Kullback relative information performed consistently better than the histogram intersection (L_1) metric, and roughly the same as L_q.

5. Similarity Noise in Stereo Matching

Stereo matching is the process of finding correspondences between entities in images with overlapping scene content. The images are typically taken from cameras at different viewpoints which implies that the intensity of corresponding pixels may not be the same.

In the first experiments we used two standard stereo data sets (Castle set and Tower set) provided by Carnegie Mellon University. These datasets contain multiple images of static scenes with accurate information about object locations in 3D. The 3D locations are given in X-Y-Z coordinates with a simple text description (at best accurate to 0.3 mm) and the corresponding image coordinates (the ground truth) are provided for all eleven images taken for each scene. For each image there are provided 28 points as ground truth in the Castle set and 18 points in the Tower set.

Let $\mathcal{I}_1$ and $\mathcal{I}_2$ represent intensities in two templates i.e. there exist n tuples $(\mathcal{I}_1^1, \mathcal{I}_2^1), \cdots, (\mathcal{I}_1^n, \mathcal{I}_2^n)$, n depending on the size of the template used. The quantity $SSD = \sum_{i=1}^{n} (\mathcal{I}_1^i - \mathcal{I}_2^i)^2$ measures the squared Euclidean distance between $(\mathcal{I}_1, \mathcal{I}_2)$ and a value close to zero indicates a strong match. The other metrics L_1 and L_c can be defined similarly.

In each image we considered the templates around points which were given by the ground truth. We wanted to find the model for the real noise distribution which assured the best accuracy in finding the corresponding templates in the other image. As a measure of performance we computed the accuracy of finding

the corresponding points in the neighborhood of one pixel around the points provided by the test set. In searching for the corresponding pixel, we examined a band of height 7 pixels and width equal to the image dimension centered at the row coordinate of the pixel provided by the test set.

In this application we used a template size of n=25, i.e. a 5x5 window around the central point. For the training sets, we placed templates around 10 points which were obtained from the ground truth.

As one can see from Table 3 the Cauchy distribution had the best fit to the real noise distribution relative to L_1 and L_2. Therefore, one expects the accuracy to be the greatest when using L_c (Table 4). In all cases the results obtained with L_2 are the worst. Furthermore, L_c has the best accuracy relative to the other analytic similarity measures for both test sets.

Set	Gauss	Exponential	Cauchy
Castle	0.0486	0.0286	0.0246
Tower	0.049	0.045	0.043

Table 3. The approximation error for the corresponding point noise distribution in stereo matching for three distribution models

Set	L_2	L_1	K	L_c	ML
Castle	91.05	92.43	92.12	93.71 $\mathbf{a}$=7.47	94.52
Tower	91.11	93.32	92.84	94.26 $\mathbf{a}$=5.23	95.07

Table 4. The accuracy of the template based stereo matcher (%)

In addition, we investigated the influence of similarity noise using two promising stereo algorithms and another stereo pair from the research literature. Our intention was to try other distance measures than SSD (which was used in the original algorithms) in calculating the disparity map.

The first algorithm introduced by Fusiello, et al. [6], is an adaptive, multi-window scheme using left-
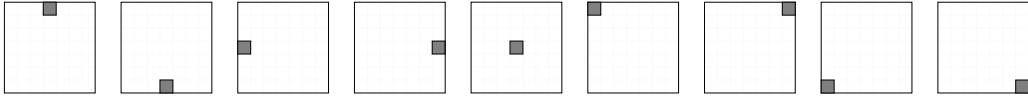
Figure 6. The nine asymmetric correlation windows

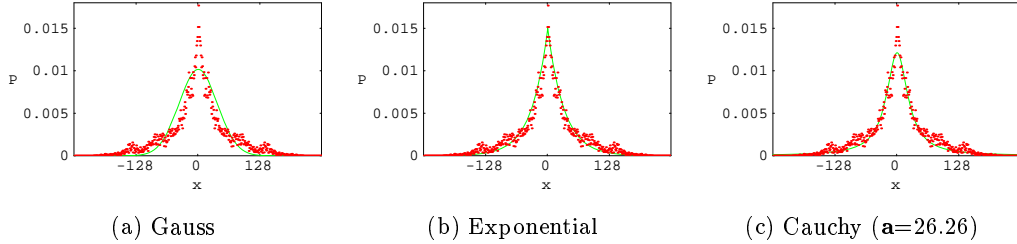(a) Gauss (b) Exponential (c) Cauchy ($\mathbf{a}=26.26$)

Figure 7. Noise distribution for the Robots stereo pair compared with the best fit Gaussian (a) (approximation error is 0.0267), best fit Exponential (b) (approximation error is 0.0156) and best fit Cauchy (c) (approximation error is 0.0147)

right consistency to compute disparity. For each pixel the correlation with nine different windows (Figure 6) is performed and the disparity with the smallest SSD (L_2) error value is retained. The authors conclude that the adaptive, multi-window scheme clearly outperforms fixed window schemes. Moreover, the left-right consistency check proves to be effective in eliminating false matches and identifying occluded regions.

The second algorithm we implemented and tested was introduced by Cox, et al. [5]. Their algorithm optimizes a maximum likelihood cost function. This function assumes that corresponding features in the left and right images are normally distributed about a common true value and consists of a weighted squared error term if two features are matched or a (fixed) cost if a feature is determined to be occluded. Their idea was to perform matching on the individual pixel intensity, instead of using an adaptive window as in the area-based correlation methods.

In order to evaluate the performance of the stereo matching algorithms under difficult matching conditions we also used the Robots stereo pair [13]. This stereo pair is more difficult due to varying levels of depth and occlusions (Figure 8). This fact is illustrated in the shape of the real noise distribution (Figure 7). Note that the distribution in this case has wider spread and is less smooth. For this stereo pair, the ground truth consisted of 1276 point pairs, given with one pixel accuracy.

Consider a point in the left image given by the ground truth. The displacement of the corresponding point position in the right image is given by the disparity map. The accuracy is given by the percentage of pixels in the test set which are matched correctly by
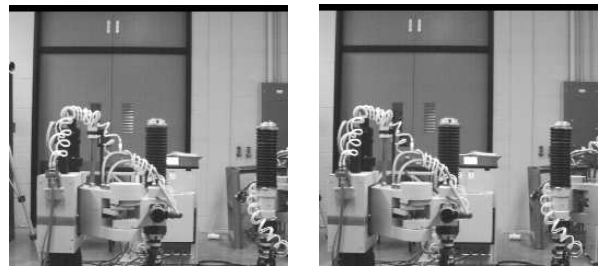

Figure 8. Robots stereo pair

the algorithm.

In Tables 5 and 6 the results using different distance measures are presented for Fusiello's and Cox's algorithms, respectively. Using ML gave an improvement in accuracy of 3 to 9 percent over the original implementations which used L_2. Among the analytic metrics, L_c consistently had the best accuracy.

Set	L_2	L_1	K	L_c	ML
Castle	92.27	92.92	92.76	94.82 $\mathbf{a}$=7.47	95.73
Tower	91.79	93.67	93.14	95.28 $\mathbf{a}$=5.23	96.05
Robots	72.15	73.74	75.87	77.69 $\mathbf{a}$=26.2	79.54

Table 5. Accuracy (%) using the multiple window stereo algorithm (Fusiello)

Set	L_2	L_1	K	L_c	ML
Castle	93.45	94.72	94.53	95.72 $\mathbf{a}$=7.47	96.37
Tower	93.18	95.07	94.74	96.18 $\mathbf{a}$=5.23	97.04
Robots	74.81	76.76	78.15	82.51 $\mathbf{a}$=26.2	84.38

Table 6. Accuracy (%) using the maximum likelihood stereo algorithm (Cox)

6. Discussion and Conclusions

In summary, we examined two topic areas from computer vision which were content based retrieval and stereo matching. Regarding content based retrieval, the first application was finding copies of images which had been printed and then scanned. For this application we used the Corel stock photo database and a color histogram method for finding the copies. The second application dealt with object recognition using color invariance. Both the ground truth and the algorithm came from the work by Gevers, et al. [7]. Note that in their work, they used the SAD metric. Furthermore, for both applications, we implemented Hafner's [8] quadratic perceptual similarity measure as a benchmark.

The second topic area we examined was stereo matching. We implemented a template matching algorithm, an adaptive, multi-window algorithm by Fusiello [6], and a maximum likelihood method using pixel intensities by Cox, et al. [5]. Note that the SSD was used in the work by Fusiello [6] and in the work by Cox [5].

For both topic areas and applications in our experiments, the ML metric consistently outperformed all of the analytic metrics. Minimizing the ML metric is optimal with respect to maximizing the likelihood of the difference between image elements when the real noise distribution is representative. Therefore, the breaking points occur when there is no ground truth, or when the ground truth is not representative.

The first problem that this paper addresses is whether the SSD is appropriate to use for computer vision applications in content based retrieval and stereo matching. From our experiments, the SSD is typically not justified because the real noise distribution is not Gaussian.

There appear to be two methods of applying maximum likelihood toward improving the accuracy of matching algorithms. The first method recommends altering the images so that the measured noise distribution is closer to the Gaussian and then using the SSD. The second method is to find a metric which has a distribution which is close to the real noise distribution. Our experiments suggest that real noise distributions can be modeled using the Cauchy distribution better than with the Gaussian or Exponential. We used the Chi-square test as the measure of fit between the distributions, and found in our experiments that it served as a reliable indicator for distribution selection. Furthermore, the Kullback relative information also appears to be more accurate in our experiments than the SSD, but not as accurate as the Cauchy metric. Either method has the potential to improve the accuracy of a wide range of vision algorithms (particularly those in which the SSD or SAD are used).

Therefore, our main contributions are in showing that the prevalent Gaussian distribution assumption is often invalid, and in proposing the Cauchy metric as an alternative to both the SAD and Kullback relative information. In the case where representative ground truth can be obtained for an application, we provided a method for selecting the appropriate metric. Furthermore, we explained how to create a maximum likelihood metric based on the real noise distribution, and in our experiments we found that it consistently outperformed all of the analytic metrics.

In future work we intend to examine the influence of multi-parameter distributions towards achieving a better fit to the real distribution.

References

[1] H. Akaike. Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory, Armenia*, 1971.

[2] S. Barnard and M. Fischler. Computational stereo, comp. survey. *Science*, 194:283–287, 1976.

[3] D.N. Bhat and S.K. Nayar. Ordinal measures for image correspondence. *IEEE Trans. on PAMI*, 20(4):415–423, 1998.

[4] R. Boie and I. Cox. An analysis of camera noise. *IEEE Trans. on PAMI*, 14(6):671–674, 1992.

[5] I. Cox, S. Hingorani, and S. Rao. A maximum likelihood stereo algorithm. *CVIU*, 63(3):542–567, 1996.

[6] A. Fusiello, V. Roberto, and E. Trucco. Efficient stereo with multiple windowing. *CVPR*, pages 858–863, 1997.

[7] T. Gevers and A. Smeulders. Color-based object recognition. *Pattern Recognition*, 32(3):453–464, 1999.

[8] J. Hafner, H. Sawhney, W. Equitz, M. Flickner, and W. Nyblack. Efficient color histogram indexing for quadratic form distance functions. *IEEE Trans. on PAMI*, 17(7):729–736, 1995.

[9] F.R. Hamper, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistic: The Approach Based on Influence Functions*. John Wiley & Sons, 1986.

[10] R. Haralick and L. Shapiro. *Computer and Robot Vision II*. Addison-Wesley, 1993.

[11] P.J. Huber. *Robust Statistic*. NewYork: Wiley, 1981.

[12] S. Kullback. *Information theory and statistics*. Dover Publications, 1968.

[13] M. Lew, T. Huang, and K. Wong. Learning and feature selection in stereo matching. *IEEE Trans. on PAMI*, 16(9):869–882, 1994.

[14] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.

[15] M.J. Swain and D.H. Ballard. Color indexing. *IJCV*, 7(1):11–32, 1991.